Project 1

Evan Cheng

I. Introduction

Marathons, a challenge to the limits of human endurance that that fascinates millions of athletes worldwide.Runners not only view marathons as a personal test of fitness but also as significant community events. However, the organization of such events requires a significant amount of planning, including managing road closures, providing adequate medical services, and ensuring the safety of both participants and spectators. Environmental factors such as weather conditions also play a important role, affecting both the planning and running of the marathon.

A critical aspect in marathon is the understanding of performance dynamics under various environmental conditions. Prior research has demonstrated that environmental factors, such as the Wet Bulb Globe Temperature (WBGT), significantly affect individual's marathon performance. Ely et al. (2007) found a slowdown in marathon performances with increasing wet bulb globe temperature (WBGT). Additionally, Vihma (2010) suggested that the effects of weather conditions are less significant in female runners. These observations invite further investigation into gender-specific responses to environmental stressors. As the popularity of marathons increases exponentially over the years, it's necessary to further validate the factors that can critically impact individual's performance.

The motivation for this study is to understand the impact factors on marathon performance comprehensively. This knowledge is essential for runners to train more effectively and for organizers to monitor participants' potential medical risks. The implication of these findings extend beyond individual races, potentially on influencing individuals' long-term health and engagement with the sport.

The primary aim of this paper is to validate existing research regarding the environmental impacts on marathon performance. This analysis report will examine how increasing age affects marathon performance in both men and women, and how weather factors influence performance across different age groups and genders. Specifically, it aims to identify the weather parameters that have the most significant impact on performance.

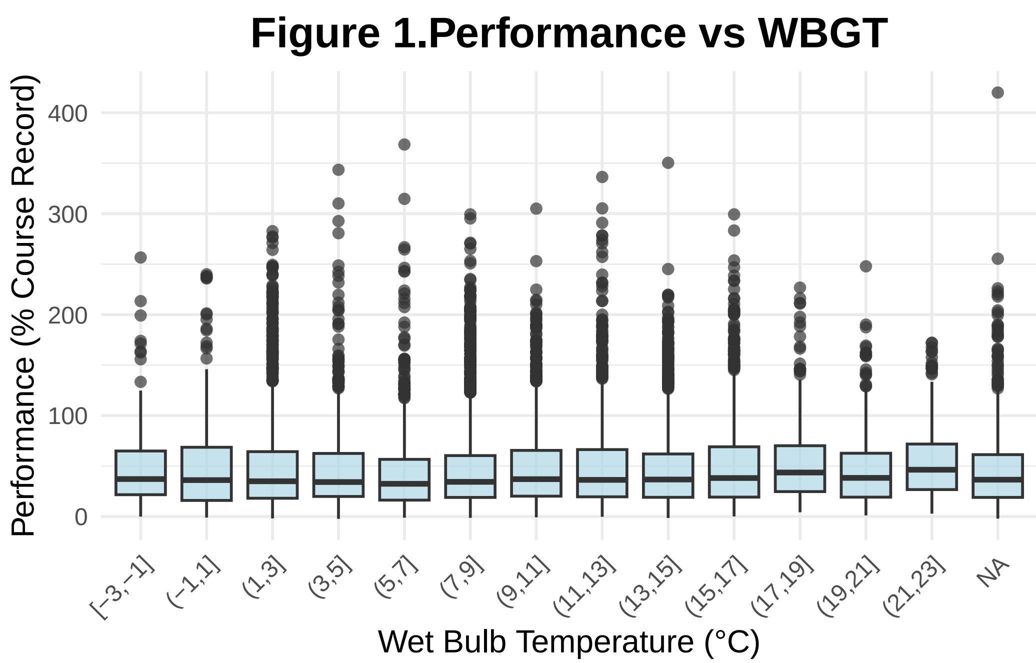
II. Data Overview

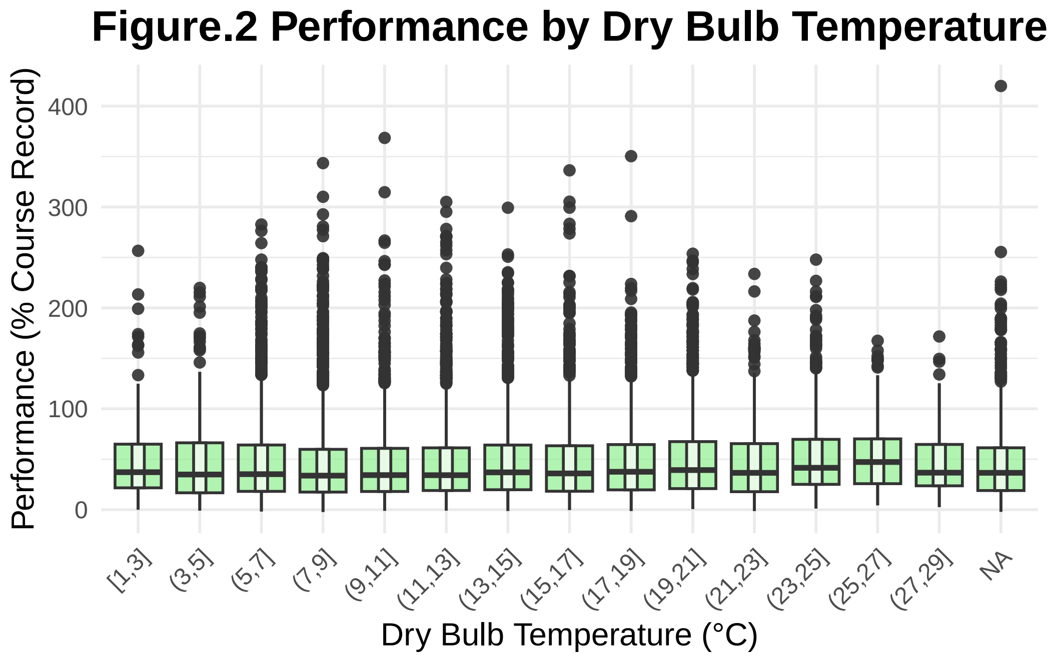
Data were collected from five major marathons: the Boston Marathon, Chicago Marathon, New York City Marathon, Twin Cities Marathon, and Grandma's Marathon. The dataset covers the years 1993 to 2016, offering more than two decades of race data for comprehensive analysis. Each dataset includes key variables related to marathon performance, environmental conditions, and participant demographics. These variables are critical for understanding how individual and environmental factors collectively influence marathon performance. A detailed explanation of the key variables is provided below.

Key Variables:

1. **Flag**: This represents the risk level for heat illness based on WBGT (Wet Bulb Globe Temperature). Categories are:
   - White = WBGT <10°C
   - Green = WBGT 10-18°C
   - Yellow = WBGT >18-23°C
   - Red = WBGT >23-28°C
   - Black = WBGT >28°C
2. **Age (yr)**: The age of the participant at the time of the marathon, in years.
3. **%CR**: This shows the percentage off from the current course record for the participant's gender. In other words, lower percentage indicates better performance.
4. **Td, C**: The dry bulb temperature, measured in Celsius, representing the air temperature.
5. **Tw, C**: The wet bulb temperature, measured in Celsius, which factors in humidity.
6. **%rh**: The percentage of relative humidity.
7. **Tg, C**: The black globe temperature in Celsius, which factors in solar radiation.
8. **SR W/m2**: Solar radiation, measured in Watts per square meter.
9. **DP**: The dew point temperature, measured in Celsius.
10. **Wind**: Wind speed, measured in kilometers per hour.
11. **WBGT**: Wet Bulb Globe Temperature, a composite measure that accounts for temperature, humidity, and solar radiation, used to estimate heat stress.

Before the analysis, an exploratory data overview was conducted, revealing several important trends and correlations between weather variables and marathon performance. Figure 1, which plots performance against Wet Bulb Globe Temperature (WBGT), shows a clear decline in performance as WBGT increases. The effect is especially noticeable above 15°C. This supports the existing findings that higher WBGT values, indicating more stressful environmental conditions, can reduce athletic performance. Additionally, the relationship between performance and dry bulb temperature further supports the negative correlation between temperature and performance (Figure 2. Performance by Dry Bulb Temperature).





From Figure 3 and Figure 4, the plots examine the performance against relative humidity and solar radiation. The results suggest that extreme humidity slightly reduces performance. The data points are scattered, suggesting varied individual reactions to humidity. Additionally, higher levels of solar radiation seem to correlate with decreased performance; it's likely due to the impact of heat from direc solar.

Figure 3Performance vs Humidity

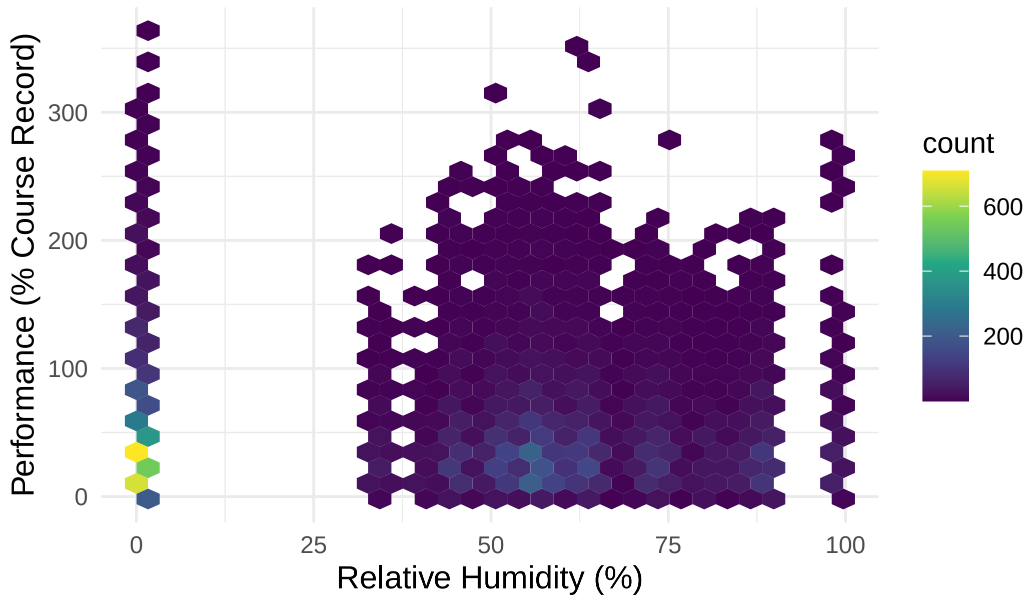
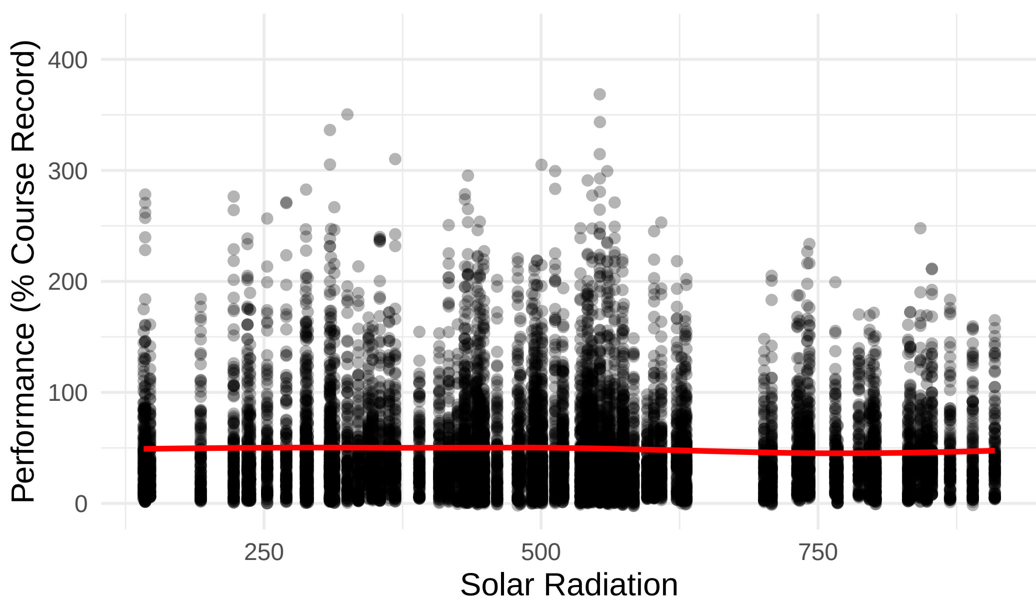
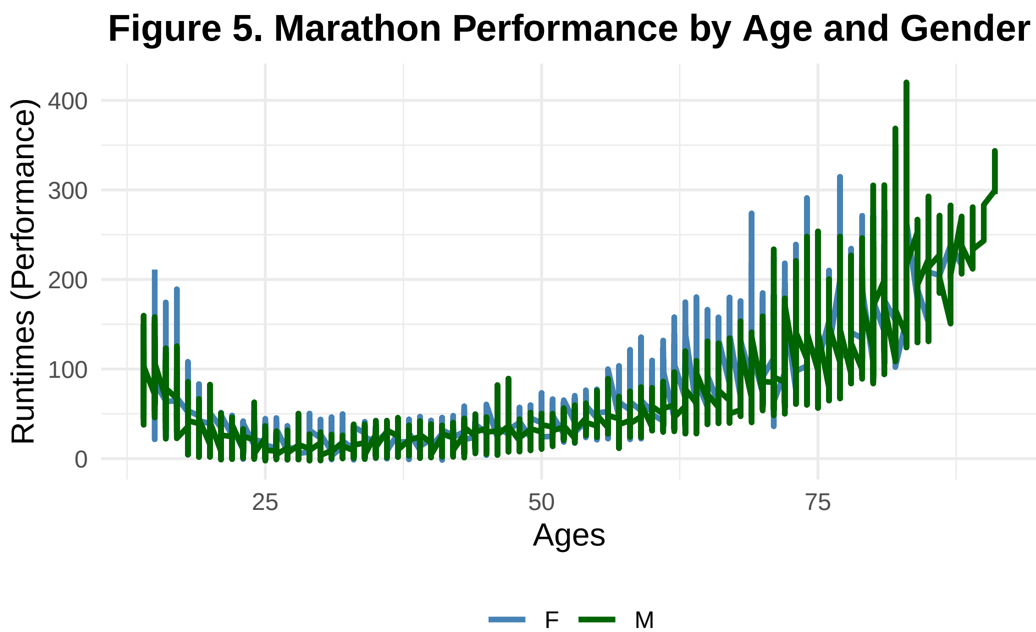


Figure 4.Performance vs Solar Radiation



Moreover, the plot comparing performance versus age by gender (Figure 5. Effects of Age on Marathon Performance in Men and Women) reveals a clear U-shaped curve. This suggests the peak age for performance is between 25 and 50. This trend appears more pronounced among male runners compared to female runners. The overall density plot of all participants' performance (Figure 5. Effects of Age on Marathon Performance in Men and Women) shows a bimodal distribution, indicating two distinct groups within the marathon participants: highly competitive runners and more casual participants. These observations underscore the need to carefully select the analysis method to accurately evaluate the relationship between the interested variable and performance.



Missing Data

The Missing Data Table reveals a 4.2% missing rate across the weather variables. It's likely that one of the years is missing all weather data. To ensure data quality, all rows with missing weather variables will be excluded from the analyses, as the data are not missing at random, and the sample size is sufficient for appropriate analysis. It's also worth noting that the current dataset only consist performances from individuals who completed each race.

III. Data Analysis

Aim 1: Effects Of Age On Performance In Men And Women

A mixed-effects model was used to examine the effect of age in marathon performance between men and women, and account for the random effect of different marathon course at different year. The model indicated several significant effects:

Table 1. Missing Data Summary

Variables	Missing Values	Percentage Missing (%)
Flag	491	4.25
Dry bulb Temp C	491	4.25
Wet bulb Temp C	491	4.25
Percent Relative Humidity	491	4.25
Black Globe Temp C	491	4.25
Solar Radiation	491	4.25
Dew Point in C	491	4.25
Wind Speed	491	4.25
Wet Bulb Globe Temp	491	4.25

Age

The analysis showed a statistically significant positive effect of age on marathon performance, with an estimate of 1.70, a standard error of 0.03, and a t-value of 65.55. In this context, a positive coefficient for age indicates that as runner grow older, their performance tends to decline. In other words, younger runners are closer to record performances, potentially due to better physical conditioning.

Gender

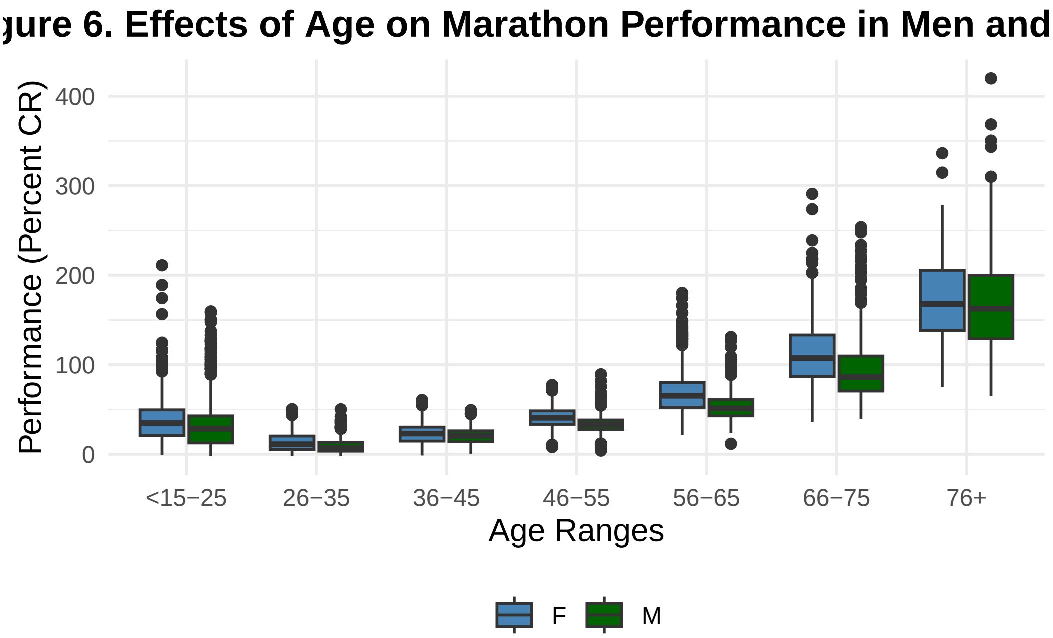
The model revealed that male atheletes are associated with worse performance compared to females, with an estimate of -9.99 and a standard error of 1.68, resulting in a t-value of -5.96. This finding suggests that males athletes, on average, are further away from the record times than females when controlling age and courses. This could be caused by a variety of physiological and training factors between genders.

Age-Gender Interaction

The interaction between age and gender was also significant, with an estimate of 0.11 and a standard error of 0.03, resulting in a t-value of 3.319. This suggests that the rate at which performance declines with age is more significant in male athletes than in female athletes. In other words, older males may experience a stronger decline in performance, this potentilly reflect differences in endurance with age for males and females.

Summary

The results highlights that both age and gender are crucial in understanding the marathon performance differences. Younger runners and females are generally closer to achieving record time, while older males show decline in performance.



Aim 2: Impact of Environmental Conditions

To investigate the impact of weather variables on marathon performance and determine if the effects vary across different ages and genders, a multiple linear regression model was applied to the cleaned data. Specifically, the cleaned dataset excluded all rows with missing data, as well as participants under the age of 18, to eliminate unnecessary noise.

This model will quantify the influence of each weather variable such as WBGT, relative humidity, and solar radiation, while accounting for age and gender. Interactions in the model will determine if and how these effects impact across different ages and between genders.

WBGT (Estimate = 0.2797, SE = 1.594e-01, p = 0.07931)

WBGT showed a positive, but not statistically significant, correlation with performance measure, suggesting a tendency for increased times (worse performance) under higher temperatures. This finding aligns with expectations regarding heat stress impacts from prior studies.

Relative Humidity (Estimate = 1.704e-01, SE = 2.963e-02, p < 2e-16)

The analysis on the impact of relative humidity on marathon performance demonstrates a clear and significant negative effect, where increased humidity levels lead to poorer performance outcomes. The model provides a estimate of 0.1704 for relative humidity, with a standard error of 0.02963, and a significant p-value ($p < 2e\text{-}16$). This positive estimate indicates that as relative humidity rises, the performance of runners decline significantly. The adverse effect of higher humidity on marathon performance can be explained by the increased physiological burden it places on runners.

Solar Radiation (Estimate = 3.22e-02, SE = 5.398e-03, p < 2e-16)

The analysis of solar radiation's impact on marathon performance reveals a significant detrimental effect, where an increase in solar radiation corresponds to slower marathon times. Specifically, the regression model indicates a positive estimate of 0.0322 for solar radiation, with a standard error of 0.005398, and a highly significant p-value ($p < 2e-16$). This statistical finding confirms that as solar radiation increases, marathon runners tend to exhibit worse performance.

Age (Estimate = 2.579e+00, SE = 6.615e-02, p < 2e-16): Age showed Age demonstrate a significant positive effect, which indicates that performance declines with age, with older runners taking longer to complete each race.

Gender (Male) (Estimate = -7.546e+00, SE = 2.365e+00, p = 0.00143): Gender in this case revealed a negative effect. That is, male athletes have significant lower performance measure compared to females, suggesting that males generally perform better than females in marathons.

Interaction Effects

WBGT:Age (Estimate = 5.218e-03, SE = 3.077e-03, p = 0.08992) WBGT:Sex (Male) (Estimate = -1.37e-01, SE = 1.071e-01, p = 0.20057) The interaction between WBGT and age and WBGT and sex were not statistically significant, which indicates that the effect of Wet Bulb Globe Temperature (WBGT) on marathon performance does not seem to vary across different age groups or genders. Although WBGT is known to impact performance negatively, these results suggest that its influence is relatively uniform across ages and genders in this dataset.

Relative Humidity:Age (Estimate = -4.683e-03, SE = 5.646e-04, p < 2e-16):

In contrast, the interaction between relative humidity and age was statistically significant. This interaction reveals that the adverse effects of increasing humidity on marathon performance intensify as runners age. The negative estimate indicates that older athletes are more vulnerable to humidity's effects, potentially due to reduced thermoregulation efficiency.

Solar Radiation:Age (Estimate = -9.514e-04, SE = 1.034e-04, p < 2e-16):

Similarly, the interaction between solar radiation and age was also significant. This finding suggests that older runners are more negatively affected by increased solar radiation. As runners age, their ability to manage the effects of solar exposure may decrease. This can lead to slower performance times in high-radiation environments.

Summary

The multiple linear regression analysis reveals serval ways in which environmental factors impact marathon performance across different age group and gender. While WBGT and its interactions show trends that suggest a negative impact on performance, the effects of humidity and solar radiation are more statistically significant. These insights are important for marathon training and planning. All results suggests that older runners and those running in more humid or sunny conditions may need additional assistance or planning to optimize their performance.

Aim 3: Influence of Specific Weather Parameters

The multiple linear regression model was applied to assess the influence of various weather parameters on marathon performance, with the ultimate goal of identifying which weather variable had the greatest impact. All variables were standardized before the model fit. Below is a summary of the key findings from this analysis.

Significant Variables:

FlagRed (Estimate = 9.702935, SE = 3.20265, p = 0.00250)

The analysis shows that the Red Flag condition, which represents extreme heat, has a significant impact on marathon performance. Under Red Flag conditions, runners experience a substantial increase in their completion times; further supporting the idea that extreme heat lead to decline in performance. This result highlights the critical influence of high-temperature conditions on endurance that can lead to increased longer race completion times.

Dry Bulb Temperature (Estimate = 1.712328, SE = 0.620774, p = 0.00582)

Dry bulb temperature, which measures the ambient air temperature, has a significant effect on marathon performance. With an estimate of 1.712328 and a standard error of 0.620774, this effect is statistically significant ($p = 0.00582$). The results indicate that as ambient temperature rises, marathon runners tend to perform worse. Again, the results support previous findings.

Relative Humidity (X.rh) (Estimate = -0.056800, SE = 0.019767, p = 0.00407)

Relative humidity also significantly impacts marathon performance. As humidity levels increase, performance times worsen. This is likely because high humidity reduce the body's ability to export heat through sweating. The negative estimate indicates that runners are slower in more humid conditions, which can be explained by heat stress.

Solar Radiation (SR.W.m2) (Estimate = -0.023866, SE = 0.003654, p < 2e-16)

Solar radiation has a highly significant effect on marathon performance. This means that higher levels of solar radiation, or direct sunlight exposure, will result in slower marathon completion times. The negative coefficient reflects that increased solar radiation adds to the body's heat load, which cause difficulty for runners to maintain optimal performance. Solar radiation is highly correlated with temperature, which supports the other findings

Summary

The findings from the multiple linear model suggest that Red Flag conditions, dry bulb temperature, relative humidity, and solar radiation are the weather parameters with the most significant impact on marathon performance. All of the above suggest that high temperatures contribute the most to slower marathon times, while higher humidity also impact the body's ability to dissipate heat, further reducing performance. Interestingly, WBGT, which combines several environmental factors, was not significant in this model, suggesting that the multiple linear regression model may not be the most suitable method for this analysis.

References

Ely, M. R., Cheuvront, S., Roberts, W. O., & Montain, S. J. (2007). Impact of weather on marathon-running performance. *Medicine & Science in Sports & Exercise, 39*(3), 487-493.

Vihma, T. (2010). Effects of weather on the performance of marathon runners. *International Journal of Biometeorology, 54*(3), 297-306.

Code Appendix

Github :https://github.com/tent284991/2550/tree/main/Project%201%20